DOCUMENT RESUME

ED 202 894 TM 810 319

AUTHOR Holley, Freda M.; Hickman, Randall C.
TITLE Research on Teacher Evaluation: Needs and Realities.
INSTITUTION Texas Univ., Austin. Research and Development Center for Teacher Education.
SPONS AGENCY National Inst. of Education (ED), Washington, D.C.
PUB DATE Apr 81
CONTRACT OB-NIE-G-80-0116
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Evaluation Methods; *Research Needs; *Teacher Evaluation
IDENTIFIERS *Evaluation Problems; *Evaluation Research

ABSTRACT

Two major functions of teacher evaluation documentation are to serve to help rationalize the process it documents, and to serve an evidential role, justifying decisions concerning personnel, policy decisions, etc. In sum, teacher evaluation documents provide access to teacher evaluation practice, the notion of teacher competency behind it, and some effects of the institutional and bureaucratic context in which teacher evaluation takes place. The materials for this analysis comprise separate teacher evaluation documents, all originating since 1974, from 12 large school districts across the nation. From the 50 documents a general picture of teacher evaluation procedures emerges which has the following features: (1) an effort to consult with the teacher and keep the teacher informed concerning the process and its results; (2) classroom observation is a ubiquitous feature; (3) diagnosis, counseling, and the offering of assistance take place; (4) a final judgment is reached based upon the relevant data; and (5) alongside the formal evaluation process, there is provision for an informal evaluation process which can give a large measure of discretionary power. Document features reflecting system needs are outlined, along with the most urgent teacher evaluation needs for research. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED202894

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

RESEARCH ON TEACHER EVALUATION:

NEEDS AND REALITIES

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
F. M. Holley
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Freda M. Holley
Austin Independent School District

Randall C. Hickman
Research and Development Center for Teacher Education
The University of Texas at Austin

Paper presented at the annual meeting of
the American Educational Research
Association, Los Angeles, April, 1981

Publication #80.55

TM 810 375

This study was supported in part by The National Institute of Education, Contract OB-NIE-G-80-0116, Classroom Organization and Effective Teaching Project, The Research and Development Center for Teacher Education, The University of Texas at Austin. The opinions expressed herein do not necessarily reflect the position or policy of The National Institute of Education and no official endorsement by that office should be inferred.

RESEARCH ON TEACHER EVALUATION: NEEDS AND REALITIES

Freda M. Holley
Austin Independent School District

Randall C. Hickman
Research and Development Center for Teacher Education
The University of Texas at Austin

Publicity and public concern about teacher competence have both focused in the last several years on teacher testing as a vehicle for improving the level of teacher performance in the nation's schools. However, one thing that has become obvious is that testing is a political issue and is unlikely to occur other than at recruitment stages. Therefore, this trend has brought greater attention to and concern about current practices in the field of teacher evaluation. Administrators have become painfully aware of inadequacies of current evaluation systems. Such focus is already spurring considerable attention to research needs in teacher evaluation.

Congruent to these developments, research on teacher effectiveness has made both evaluators and researchers aware of the urgency and importance of certain teacher competencies in the production of learning outcomes. This importance in turn also highlights to public school administrators the importance and necessity for evaluating teacher performance with respect to these competencies. Again, however, administrators find that current methods of teacher evaluation are inadequate to the task.

A third development that converges with these trends to create a favorable climate for research on teacher evaluation is the rising interest of teachers unions or associations in assuring that teacher evaluation practice is fair and equitable to teacher members. This coupled with the political power teachers are gaining through such organizations means that current teacher evaluation practices must either be improved or abandoned. Legal issues are intricately involved.

These events coalesce to suggest that research on teacher evaluation has considerable priority on the national research agenda for education. Few researchers or even evaluators working in public school settings, however, have worked intimately with teacher evaluation. Therefore, there is considerable likelihood that the first wave of research in this area will be characterized by naiveté.

The objectives of this paper are to draw on the experiences of an evaluator who has been deeply involved in teacher evaluation, an awareness of the extant literature on teacher evaluation practice, a set of teacher evaluation forms and procedures collected from large school districts

nationally in order to present a coherent statement on the needs for research on teacher evaluation, and the data yielded from attempts to improve a school district evaluation system in order to present a picture of the research needs.

What Are the Realities of the Teacher Evaluation Process?

Researchers who might wish to provide information that would assist administrators in the design of improved evaluation procedures must first have a good picture of the way it "really" is for the typical persons charged with conducting teacher evaluations. Let us first examine two illustrations of those typical persons representing extremes of the types found. Unfortunately, both extremes are likely to exist in the same system so that any process that is designed must be flexible enough to meet both limitations and to realize potential.

Semantha Richards is a principal in Stone Elementary, a school with an enrollment of about 330 students. The school neighborhood is fairly stable and teacher turnover also is very low. There are two teachers per grade in K through 4 and one each for grades five and six; there is also one class of combined fifth and sixth graders. The school has a resident librarian, a music teacher, and a P. E. teacher; there are four special education teachers because the school serves a center for hearing impaired students. The school also has a secretary and two teacher aides. Miss Richards taught school for 8 years prior to becoming a language arts supervisor for the district. She served in that role six years, then gained her administrative training and was appointed principal. In each of these roles, Semantha Richards was a warm, capable leader. Miss Richards saw evaluation as a natural part of her supervisory role. Since she has expertise in almost all those areas where she evaluates performance, her employees respect her judgments and react well to her staff development recommendations.

John Simpson is the principal of Wilson Senior High School where the enrollment currently stands at 3,340 students. There are approximately 250 faculty members. The school has both a program for gifted science students and a vocational training program keyed to the regional area's principal industry, electronics. John Simpson was a football coach for ten years before gaining his administrative certification and moving into a junior high principal's position. He served there for six years and four years as assistant principal at Wilson. He became principal there the previous year when the former principal retired. This occurred at the same time that the school district underwent a major court order involving cross-town busing for students at Wilson and the other high school in town. Simpson is feeling considerable pressure from the school's parents and from the rather tense student relations in the newly integrated school. Also, many of the faculty members have been involved in the transfer and are making unfavorable comparisons in public between Simpson and the principal in the other high school. Thus, in the evaluation situation Simpson must be

very careful about his judgments. His staff is all too aware of his lack of background in some of their specialized class offerings which means he must place undue reliance on the recommendations of the district's curriculum supervisors.

These two administrators are fictional, but the two extremes they represent are all too real. Of course, every point between these extremes will also exist.

A school district in devising a workable personnel evaluation system must recognize such extremes. The system must enable even the John Simpson's to carry out the mandated policy requirements. Thus, an evaluation system that any organization designs must meet its first reality test by assuring that the system can enable all administrators, regardless of their skills or status, to meet their policy obligations.

The system must equally assure that competent administrators such as Semantha Richards will have the freedom to exercise their full capabilities in supervision. For example, if the system requires that teachers be evaluated when their contract is up for renewal, it probably must also specify that the administrator may carry out evaluation at any time even *when contracts* are not up for *renewal*. If the evaluation system specifies 52 performance categories to be evaluated, it must also indicate that other areas not specified are also open for evaluation.

The system must equally assure that an evaluator faced with 100 personnel to evaluate in one year can cope with the task. For example, it may be necessary to enable the principal to use a team approach where assistant principals or other school administrators are authorized to carry out evaluation activities.

The evaluation system must be able to withstand almost constant criticism from the unions or professional associations of both the evaluatee and the evaluator. It must also be capable of withstanding the legal challenges that are sure to come at some point.

In general, the theoretical purposes served by a school system teacher evaluation process are the communication of system expectations and the improvement of teacher performance. While these purposes are those which serve the greatest number of teachers in the system, it must also assist the district to terminate a very limited number of staff members who cannot reach minimum performance levels even with considerable assistance. Impossible? Perhaps. It could be that devising evaluation systems that meet the ideal is impossible. Nonetheless, school districts must and will have evaluation systems that try to meet these diverse demands. Researchers should accept the challenge of helping them devise systems that are as good as possible.

Realities of Teacher Evaluation Documentation

If we look at documents being used in some of the nation's major school districts, what can we learn? Here in documentation and other associated supporting materials are valuable sources of information for teacher evaluation. We can see this when we look at the purpose of the evaluation documents.

Speaking most generally, their purpose is to provide some order for the process of teacher evaluation; in other words, their major purpose is to rationalize that process. There are many things included in this rationalization, some of which are: recording of information of various kinds (the information pertinent to assessing teacher competency being organized in a number of categories or descriptions); recording assessments of some of that information (according to certain criteria, implicit or explicit); recording an overall or summary assessment of teacher competency; supplying specific evidence, justifying certain judgments, on occasions; supporting diagnostic efforts and efforts to improve teaching; documenting the evaluation process; indicating responsibility for that process; providing for adjudication of disagreements; and providing permanent records. Notwithstanding the wealth of identifiable features here, almost all fall into three groups, each providing different information concerning teacher evaluation: 1) those concerned with the definition, assessment, and improvement of teacher competency; 2) those concerned with the evaluation procedures as a whole; 3) those concerned with meeting certain needs of bureaucratic institutions.

Assuming that documentation serves as a component in a rationalizing process, it follows that behind its use of documentation is the assumption that the information contained in and the categorization articulated in the relevant documents are logically relevant to the primary objective: a judgment concerning teacher competency; that is, the information and categorization--in short, the items of the relevant documents--represent an articulation or theory of teacher competency. Although the presentation of a notion or theory of teacher competency is not the primary purpose of the document, teacher evaluation documentation cannot avoid expressing such a notion or theory nevertheless. The possibility of reading evaluation documents as notions of teacher competency is a possibility which is grounded in their very nature and function as documents.

But teacher evaluation documentation can be read for more than information concerning a theory of teacher competency lying behind the articulation of the document. Teacher evaluation documents function in a system, a teacher evaluation practice, and thus can shed some light on system properties or properties of teacher evaluation practice. The documents "document" teacher evaluation practice. In addition, the evaluation practice as well as the documents which help to systematize and rationalize that practice are embedded in an institution. This fact and its impact upon teacher evaluation are strikingly evident in certain features of the documents.

The initial statement of the purpose of teacher evaluation documentation provides some initial guidance for understanding it. It designates the major domains which have their residues in the documentation and provides the central orientation in our approach in analyzing them. The documentation and associated materials provide access to three important areas in teacher evaluation. 1) They provide access to the "official" thought about evaluation and teacher competency which are the driving forces behind the evaluation practice; and 2) They provide access to evaluation practice; and 3) They provide access to bureaucratic and institutional needs which impact upon teacher evaluation and the nature of that impact. A further, more specific look at the function of documentation will provide further focus for our analysis.

There are two major functions. First, documentation serves to help rationalize the process it documents and is one of the major instruments of such rationalization, the other being the embedding of practices in institutions with defined roles. (The "rationalization" of a process, in general, refers to several features: methodical procedure, regularity over application, its quality as a terminating procedure with a definite outcome, "objectivity.") With respect to teacher evaluation, one feature of this rationalization is particularly important--stabilization of a practice such that teachers are evaluated in the same manner, subject to the same tests, etc. That is, one purpose of rationalization is to insure "fairness." However, there is another equally important sense in which a practice though the same over instances--teachers being treated in the same way by it--can nevertheless be unfair by employing measures which are not valid. Both are equally serious threats to the rationalization of teacher evaluation.

It should be pointed out, however, that the presence of documentation as a guide does not alone guarantee absolute stability for a practice. Here you face the problem afflicting rule-governed behavior in general. To put it most briefly, the same documentation is compatible with a range of practice. In other words, a given evaluation document is only a description of an evaluation practice; two evaluators may accept the same description of their practice although the actual evaluation practice in each case may be slightly different. Although the same terms may be used in the documentation, they may not have exactly the same meaning for each evaluator. To put it somewhat differently, a given evaluation document, which the evaluator fills out to document an evaluation, can be regarded as prescribing norms or rules for evaluation practice, but the evaluators may not apply those norms or rules in exactly the same way or in the way that the designer intended. In other words, there is the problem of "evaluator reliability." (It is, no doubt, probably for this reason that in most school districts, the forms which serve to document evaluations are usually accompanied by supporting materials which provide further details concerning items on the evaluation document. The intention here may be to reduce any ambiguity in the interpretation of certain items on the evaluation document by evaluators.) Thus, evaluation documents may not give us exact information concerning the practice which they document. In addition to this more or less accidental "evaluator unreliability," there is

also the problem of intentional efforts on the part of evaluators to subvert the specifically "documentary" purpose of teacher evaluation documentation. The possibility of "fudging" or "gundecking" cannot be ignored. Evaluation takes time, and an evaluator may occasionally succumb to the needs of the moment and falsify the documentation. Though there is no reason to think that this occurs more than very infrequently, getting an accurate picture of the realities of teacher evaluation requires that we be sensitive to this possibility.

These two possibilities do not undermine in any significant way the attempt to draw inferences concerning evaluation practice from evaluation documents. On the whole, the documents do provide important information concerning evaluation practice. Most evaluators take teacher evaluations seriously, and, hence exert some effort to insure that there is a fairly straightforward relationship between any evaluation documentation they submit and teacher evaluation practice. Under these circumstances, evaluation documentation can be taken as a good indicator of evaluation practice.

Second, documentation serves an evidential role, justifying decisions concerning personnel, policy decisions, etc. It thus must contain information which counts as evidence. The important question here is whether the documents are capable of supplying evidence which is adequate and appropriate. A glance at evaluation documents may dismay some education researchers on this score, but then it must be pointed out that policymakers and administrators have different perspectives. The notion of "adequate evidence" or "adequate documentation" for, or of a given judgment concerning teacher competency may vary depending upon one's perspective or interest. The question, "When has one adequately demonstrated the presence or absence of teacher competency?" may have more than one answer.

In sum, teacher evaluation documents provide access to teacher evaluation practice, the notion of teacher competency behind it, and some effects of the institutional and bureaucratic context in which teacher evaluation takes place. It is the purpose of the documents analysis to provide information conerning these areas.

The materials for this analysis comprise about 50 separate teacher evaluation documents, all originating since 1974, from 12 large school districts across the nation. These materials can be initially and roughly divided into two kinds. First, there are those materials which contain content closely related to the measurement and documentation of teacher competency. Generally, these materials either provide and define criteria for teacher competency, or record judgments of teacher competency, or record evidence relevant to judgments of teacher competency. A second group of documents are primarily directed toward the teacher evaluation process as a whole. These materials contain information concerning the stages of teacher evaluation process, the proper sequencing of these stages, the appropriate deadlines, and other bureaucratic and institutional references of teacher evaluation.

These materials can be further grouped in a way which is particularly useful for the purposes of this analysis. This further breakdown of the materials results in the following groups: 1) There are what can be called teacher evaluation forms. Within this group, three further distinctions can be made, or three subgroups can be identified. a) The forms which document the evaluation (either preliminary or summary) and record judgment (either preliminary or summary); b) forms which record information relevant to teacher evaluation (either preliminary or summary), such as classroom observation data; c) forms for diagnostic purposes or purposes of rendering assistance. 2) There are forms relevant to the formative stage of teacher evaluation--the identification of problems, areas to be worked on. 3) There are supplementary materials which provide an expanded articulation of teacher competency criteria. 4) There are supplementary materials containing information concerning or prescriptions for the evaluation process as a whole.

It is important to keep in mind that in addition to this classification, the materials are subject to another very important principle of organization. In order to correctly analyze the documents, it is essential that it be kept in mind that each document is part of a system of documents representing a teacher evaluation system for a given school district. Analysis which is concerned not only with the notion of teacher competency behind teacher evaluation practice but with the logic of teacher evaluation practice as well must be sensitive to the relations that documents in a given system have to one another. Hence, though there may be certain features and functions which are necessary for an adequate teacher evaluation system, there may be some variation across teacher evaluation systems in their organization, in their distribution among the documents (whether certain distributions are "better" than others is a very important research question). Analysis shows, however, that this variation is relatively small.

Though the materials do not permit a clear, complete construction of the teacher evaluation system for each school district, on the whole a fairly clear picture of the notion of teacher competency and of teacher evaluation practice emerges from the documents--clearer in some respect than in others.

What do the teacher evaluation documents suggest concerning the notion of teacher competency or effective teaching sedimented in current teacher evaluation practice? First of all, there is clearly a notion of teacher competency, and the evaluation is structured or focused accordingly. But more interestingly, rather than a single approach to teacher competency, the documents indicate the prevalence of a two-faceted approach with two correspondingly different evaluation "logics." The documents indicate that at least 60% had this dual approach and notion of teacher competency.* Briefly, these two approaches are: 1) the competent teacher as a goal-setter and goal-achiever, and 2) the competent teacher as a possessor of

*The analysis is based on the documents received from the school districts. There is good reason to believe that some districts did not submit all of the available relevant documentation. Consequently, the ability to make references concerning teacher evaluation in these districts is limited.

a multitude of skills or able to produce a number of features definitive or associated with "effective teaching."

Examining the former first, the following could be taken as a typical or paradigmatic formulation of the function of goal-setting for teacher evaluation: "Performance goals are intended as a focusing tool, permitting each educator to select, from among the sum of designated responsibilities which constitutes the baseline for evaluation, several specific areas for personal emphasis." (School district E) The following is also a typical statement of the domains from which the goals are to be selected: "Performance goals will be related to the following: the instructional program, other reasonable work requirements set forth in board goals, district policies and procedures; the relevant position description; and the Standards of Performance for _____Educators." (School district E)

Behind the goals approach may be the feeling that it more nearly captures what teaching is like, in contrast with the assessment of whether the teacher possesses certain skills, and that teaching is a process of selecting some instructional goals, devising means to achieve them, and then attempting to reach the goals by the effective employment of means. There may also be behind this the notion that determining whether teachers can produce student learning as an outcome might be a more meaningful approach to teacher evaluation. The presence of this approach in teacher evaluation may be the outcome of accountability schemes adopted at higher administrative levels which have filtered down, and thus the result of a general movement concerning the notion of evaluation.

However, the suitability of such an accountability scheme in some contexts--such as business or industry--does not guarantee its suitability here. Features present in this context--such as the greater likelihood of basing goal decisions on inadequate information--suggest that realistic goal-setting may be more difficult here. The recognition of this may be behind a special flexibility that we find evidence of in the documents concerning the goal-setting approach to teacher evaluation, as indicated in the following:

> "Final appraisal must be based on the accomplishment of goals ...In order to provide a standard of non-compliance, goals should be set with a minimum and maximum range of accomplishment stated. The teacher will be expected to accomplish the minimum progress stated to be evaluated a competent. Lack of accomplishment may lead to a restatement of goals in more realistic or specific terms." (School district D)

Rather than accept the results of a particular goal-based evaluation, the evaluator may, in effect, decide to do it over again, with different goals. An evaluation procedure which can lead to this is, at the least, not cost-effective. In addition, many evaluators are worried about the validity of this approach, believing that there are factors involved in the production of learning outcomes which are beyond the control of the teacher. Hence

the teacher should not be evaluated solely upon his or her ability to produce learning outcomes. It is this worry which is no doubt behind the use of the second approach to teacher evaluation and teacher competency indicated in the documents.

In the second approach to teacher competency, the evaluator is not assessing whether certain previously agreed upon objectives have been met, but whether the teacher behavior possesses the attributes definitive of the standards of effective teaching.

The assessment of teacher competency in this general way, according to criteria (without the identification of particular instructional goals) would seem to be the primary emphasis of teacher evaluation as indicated by the documents in the sample. All but one of the districts in the sample clearly employ this approach as part of the evaluation and all but one of the districts that employ the goal-setting approach also employ this approach.

A look at the documents which provide the relevant information indicates quite a bit of agreement concerning the areas of the domain covered by this notion of teacher competency, though these areas might not be related to other areas always in the same way or described in the same way. But behind the different distributions of terms on these relevant documents, basically the same areas are being identified.

The criteria for teacher competency in general--as defined by the relevant documents--can be roughly divided into two large groups. It should be pointed out, however, that this division may be a bit fuzzy at the edges. That is, there are some criteria which could be considered as falling into either group, and, of course, it should not be inferred that there are no relations between the criteria in one group and those of the other. That this division can be made is particularly interesting and provides some insight into the realities of the role of the teacher in the institution as well as teacher evaluation itself. It should also be pointed out that not all the relevant documents are structured in a way so as to make this distinction manifest; not all the documents are designed with the intention of making it clear that the criteria can be divided into two groups. Nevertheless, though there is variation in the organization of these criteria in the documents, a clear distinction can be made between two different kinds of criteria.

In one group are the criteria which can be referred to as "professional competencies" or "instructional competencies." Criteria in this group are most closely related to instruction, or the act of teaching as such, in contrast with the criteria in the other group which are generally, if at all, indirectly related. The second area of criteria, constituting the second group, is less centrally focused than the instructional area, and is more variously described. This second area is a collection of things, among which are items which could be variously called "human relations," " noninstructional duties," "other duties and responsibilities,"

"personal qualities."

Turning to the second area first, some of the criteria included here are the following: The meeting of duties and responsibilities in job description; sharing of school responsibilities; relationship with students, parents, and staff; cooperation with other teachers and staff; provision of opportunities for the development of people leadership; professional development as identified in position qualifications; adherence to work rules, work procedures, and work methods; possession of good judgment interpersonal relationships; the meeting of contractual obligations.

Apart from the question whether this area can be meaningfully circumscribed by "teacher competency," it is clear that teachers are being evaluated with respect to this area. Teacher evaluation is not simply "teaching" evaluation strictly speaking; teacher evaluation is not solely a question of teaching skills considered to lead to "the production of learning outcomes."

In the presence of this area in teacher evaluation, we can see one impact of the institutional context upon teacher evaluation. Teachers do not simply teach; they are functionaries within an institution. They must function compatibly with that institution, although the primary function or objective of the institution may be the "education" of its "clients." As a result of satisfying this objective in an institutional way, certain other needs are generated--needs not directly relevant to teaching of students. Because teachers must function within the institution, they are necessarily subject to satisfying these needs. In one school district guidelines, for example, it is noted that "this is an extremely important item."

> "While evaluation will concentrate upon selected areas for each individual, the employee will be expected to maintain effective standards of performance in all areas of assignment and responsibility....An 'effective' evaluation indicates that an evaluatee is aware of those duties and responsibilities, in addition to the basic function of the position...and is meeting these requirements and duties in a professional manner." (School district K)

The instructional area of teaching competency can be further organized into several areas. Of course, there is some variation across the school districts in the amount of discrimination or articulation within these various areas. The discriminations themselves may be grouped variously. The criteria comprised within the instructional area refer to skills and knowledge deemed directly necessary to successful teaching. Some of the criteria included here are the following: Knowledge of the subject matter; knowledge of student abilities; organization and management skills, and that includes the ability to complete tasks on time, to use time in a productive manner, to tailor teaching toward identifiable goals, and the planning of instruction in agreement with school and district policy; the ability to communicate with and motivate students; the ability to control

the learning environment, where this includes maintaining classroom discipline and responsiveness to features of the environment which may hinder learning.

As a whole, the documents suggest the following image of the "competent teacher": First, the competent teacher sets goals in the areas of his/her responsibility, rationally selects the means to achieve those goals, and is fairly successful in implementing those means and achieving the goals. But alongside this, however, competent teaching presents another visage: within the classroom of the competent teacher, we find well-organized instruction, motivation of students, maintenance of a suitable learning environment, the enforcement of student accountability, etc.; outside the classroom, the competent teacher does his/her share in meeting institutional ("housekeeping") needs.

Not only is there a notion of teacher competency embedded in the documents, a notion of the teacher evaluation process can also be elicited from the documents.

One of the first things that comes to mind is the characterization of teacher evaluation as satisfying an important system or institutional need. Here teacher evaluation is conceived as part of the need for evaluation in general, as the following indicates:

> "Appraisal must be used by the appraised and appraiser as a positive process. It must serve to aid the organization in making decisions about its management and managers. It must serve to have the manager know where he or she stands. It must give all parties greater insight into identifying areas of strength and weakness. It must suggest possibilities for greater effectiveness." (School district D)

This emphasis upon evaluation of the system in general may, no doubt, be motivated in part by certain public demands, as the following suggests: "Accountability and appraisal have come to be synonymous terms as school patrons and citizens ask the question: 'What about the school system's productivity? Is it effective? What have we received for our money?' You must have answers to such questions." (School district D) Teacher evaluation, then, is just part of the institution's self-evaluation as a whole, designed with the intention of improving institutional functioning as a whole.

But improving the efficiency or increasing the productivity of the system does not mean simply getting rid of weak personnel. Another feature common to evaluation practice suggested by these documents is the effort toward diagnosis and the provision for assistance. There is clearly an intention (at least officially expressed) not to simply get rid of a teacher who is having problems, but to use resources to assist that teacher in improving, and, in particular, to provide that teacher with the opportunity to improve before any kind of final evaluation. As

a result, preliminary and interim conferences with the teacher, and the flexibility of having additional conferences, are notable features of the evaluation practice, the documents indicating that they occur in at least 2/3 of the districts. These conferences serve more than one function. First, they can serve to identify objectives, which in some cases may be important terms of evaluations. In other words, they can serve to produce an agreement concerning what is expected of the teacher. Secondly, they can serve the communication of some general standards of teacher performance. Thirdly, they serve to maintain lines of communication with the teacher and provide the teacher with a more active role in the evaluation process. An emphasis upon consultation with the teacher and making the teacher an active, rather than a passive, part of the process is another striking characteristic of the evaluation practice, as such as suggested by the documents. Corresponding with the concern to involve the teacher, there is also a concern to agree upon methods for evaluating progress and the establishment of formalized procedures in case of disagreement between judgment of evaluator and the judgment of evaluatee.

The documents included manifest concern on the part of the school districts to portray teacher evaluation in a positive light, de-emphasizing its function to determine the hiring and firing of personnel, as well as to exert a general measure of behavioral control on personnel. In some cases this concern for a positive portrayal can extend pretty far, leading to the making of rather large claims, as the following indicates:

> "The following are the essential uses of the appraisal: Performance improvement; motivation toward achieving personal and system goals; discovery of abilities; self-development; discovery of educational/professional needs; facilitate understanding between supervisor and subordinates; determine career potential; guidelines for transfer, promotion, reassignment; guide for salary determination." (School district D)

It is certainly legitimate to wonder whether a teacher evaluation system can in any meaningful way serve all these ends.

Although the documents indicate that teacher evaluation is structured, with formalized procedures (presumably for the sake of rationalization of the process), they also suggest that in many systems there is a provision for some flexibility. "Teacher evaluation" is not only the formalized procedure, but also a much less formalized one.

> "Evaluation is a continuous process and may occur between scheduled periods at the request of the employee, designated evaluator, or the supervisor....Actions taken by administrators as a normal part of their supervisory responsibilities <u>may</u> ultimately become a part of the formal evaluation process. They may be used as a step-by-step process <u>or</u> any one of these steps may be used independently as circumstances indicate." (School district K)

This flexibility in the actual practice or behavior referred by evaluation can go quite far: "The term 'evaluation' refers to any action of the supervisor to establish standards of performance, behavior, and to hold personnel under his supervision responsible for meeting the standard." (School district K) This notion of teacher evaluation gives the evaluator a great deal of freedom in determining what counts as "evaluation," notwithstanding the desire of the school district to avoid subversion of the formalized process:

> "This statement, as to the broad scope of the total process of evaluation, does not mean that the evaluative provisions of the Education Code should be avoided or subverted, but only that the broader scope of administrative evaluation and supervision can be used to supplement the specific evaluation process. These supplementary methods may be incorporated into the district's formal performance evaluation systems as appropriate." (School district K)

The politics of teacher evaluation emerges between the lines here: In his function as evaluator, the supervisor has a general power to act vis-á-vis teachers. This notion threatens to reduce the entirety of the action of the supervisor within the institution to "evaluation." It is not difficult to imagine situations in which this possibility might be abused. This is in sharp contrast to the amount of control that teachers have over the evaluation process and its definition.

In sum, a general picture of teacher evaluation procedures emerges from the documents--a picture with areas of agreement as well as areas of variation--which has the following features:

1. There is an effort to consult with the teacher and keep the teacher informed concerning the process and its results. But although conferences between evaluator and evaluatee are an explicit feature of all but one district in the sample, there are variations concerning whether preliminary and interim conferences are formalized procedures as well as the total number of conferences.

2. The relevant data for the evaluation come from a number of sources, but classroom observation (and the use of instruments designed for this purpose) is a ubiquitous feature, with variations in the extensiveness and detail of the observation instruments.

3. Diagnosis, counseling, and the offering of assistance take place though there is variation in the documents concerning the amount of emphasis upon this.

4. A final judgment is reached, based upon the relevant data. But there is surprisingly little effort in the documentation to be explicit about this decision-making process and the questions

that inevitably arise, such as: How is the performance in the various areas to be weighed with respect to each other? The following is typical of the way this issue is treated in the documents:

> "The composite evaluation shall reflect the judgments made in Sections I and II. If some elements of Sections I and II have been evaluated to be less than 'effective,' yet other elements 'effective,' the professional judgment of the supervisor will determine whether the composite evaluation is 'unsatisfactory,' 'requires improvement,' or is 'effective.'" (School district K)

5. Alongside the formal evaluation process, there is provision for an informal evaluation process which can give a large measure of discretionary power. An important question that arises here concerns the kind of impact (of a subversive or constructive nature?) that an informal evaluation process can have on the formalized system.

This is only a sketch of the typical evaluation process, but even as a sketch there are some other features that need to be mentioned. The data which has been presented concerning the notion of teacher competency and the notion of the teacher evaluation process embedded in the documentation must be supplemented by a third area which pays particular attention to some institutional effects. Some attention has already been directed to this, and a little further reflection upon what has been presented can reveal further instances of the impact of the institutional context upon teacher evaluation at certain points. But it is worthwhile to note some features (some quite obvious, others less so) which manifest the presence of a bureaucratic institution as a context in a striking way.

Some document features which reflect system needs or the institutional context are the following:

1. Signatures (evaluator & evaluatee) are present--attesting to the validity of the documentation; indicating responsibility for the evaluation and summary judgment and the awareness of the evaluatee concerning the results; manifesting proof needs, indicating that document is the official document of results.

2. There is provision for disclaimers and comments on the part of the evaluatee when there is disagreement concerning a feature of the evaluation (primarily the summary judgment--a reflection of political and legal needs).

3. There are multiple copies (two or three) usually for the evaluator (usually principal), evaluatee, and personnel department; this bears witness to the informational needs of the institution.

4. The materials which are designed to serve a documentary purpose (which are usually multiple-copy forms) are frequently simplified, and do not record much data (if any); the main purpose is to indicate the summary judgment and the composite judgments in the other areas; and provide information to identify the evaluatee. (Other information requisite to execution of evaluation is usually located in supplementary materials which have little circulation in the system.)

5. In evaluations which indicate "unsatisfactory" (in part or whole), there is provision for justifying information (Such information is necessary.) on the summary judgment document or an addendum document. This is clearly a manifestation of political and legal needs.

6. Probationary and tenured teachers are not treated in exactly the same way by the evaluation process. There are differences not only with respect to the frequency of formal evaluation but the closeness of scrutiny as well. No doubt, at the least this reflects a need to economize evaluation efforts, but there may be political factors involved as well.

7. The same form sometimes serves for the interim as well as final evaluation.

8. There are formalized procedures in the case of disagreement between evaluator and evaluatee concerning the results or the procedures of evaluation.

The analysis of teacher evaluation by means of evaluation documents examines teacher evaluation from one of the major perspectives employed in this paper, a perspective which takes its departure point from a crucial juncture at the center of teacher evaluation--between teacher evaluation in its functional role and the institution in which it functions. Indeed, the very possibility of this kind of analysis--presupposing the existence of documentation--rests upon the existence of this juncture, and its significance for teacher evaluation should not be underestimated.

The results of this analysis raise many questions which teacher evaluation research cannot afford to ignore--questions concerning the impact of certain features on the validity of teacher evaluation practice and the notions of teacher competency; questions concerning the impact of the politics of the institutional context--not just on teacher evaluation, but also on the efforts of teachers to achieve a full professional status; questions concerning the necessary features of any adequate teacher evaluation system and how such adequacy is to be recognized.

The presence of informal, in addition to formal, teacher evaluation raises certain questions. How does the presence of informal teacher evaluation affect the formalized teacher evaluation system? Does it contribute to or hinder rationalization of the evaluation process, which is

embedded in the formalized procedure? Could it lead to an erosion of the formalized procedures? What is the significance for teacher evaluation practice of conceiving of evaluation as a "continuous process"? Answering these questions, no doubt, will at one point or another lead to an examination of the politics of the education institution and the politics of teacher evaluation, for the latter must certainly be regarded as affected by the former.

With respect to the notion of teacher competency articulated in the documents, the intrusion of institutional needs was noted. While no one would suggest that we not keep in mind that teachers are employees of an institution (unless we are willing to consider suggestions of vast changes in the way that education takes place), we must be sensitive to the possibility of a conflict of interest here. There may be instances in which satisfying institutional needs may conflict with satisfying instructional needs. Hence, it is important that the priorities of teacher evaluation be established, the relative importance of the teacher's role as a bureaucrat or institutional member versus the role of the teacher as an instructor.

Realities of the Teacher Evaluation Data

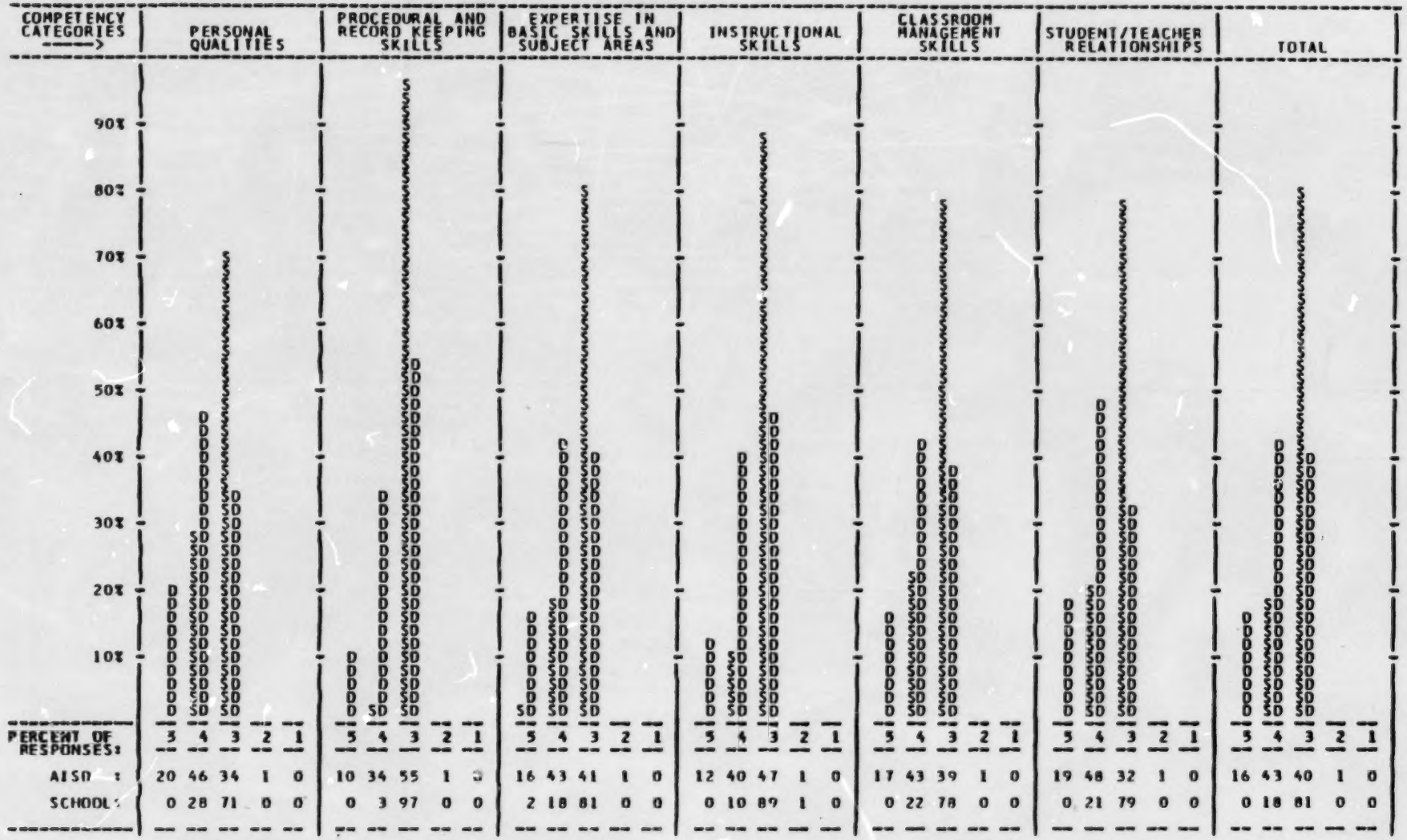
Traditionally, the teacher evaluation data is produced and stored in the teacher's evaluation folder. Unless termination is involved, it then gets considered only if the teacher's personnel file is pulled for consideration in transfer or promotion. For this reason, the characteristics inherent in this data are rarely revealed. When they are systematically gathered and analyzed, however, the realities are given! Figure 1 contains output for one school in a system which tabulates each school's complete data against the average for the entire system. It is indicative of the lack of consistency of such data across evaluators.

As it turns out, the data not only lacks such consistency, it also contains enormous bias. Analyses have revealed biases on the basis of ethnicity, sex, grade level, subject matter taught (Christner 1981).

> "Even across changing evaluation forms and systems, certain trends were noted with relatively few exceptions. Female evaluatees had higher ratings than did male evaluatees. Generally, ratings increased with an increase in teaching experience (leveling off at a certain point). Ratings were lowest for temporarily employed professionals, next lowest for probationary employees, and highest for professionals on three-year or grant contracts. A professional with a Master's degree generally had higher ratings than did a professional with a Bachelor's degree. Although there was some change in Anglo and Mexican-American evaluatees receiving the highest ratings, Blacks generally received the lowest ratings as a group. Counselors and librarians generally received higher ratings than did special education professionals, who

1979-80 SUMMARY OF TEACHER EVALUATIONS FOR LANIER HIGH SCHOOL (JULY, 1980)

COLUMNS OF LETTERS REPRESENT PERCENTAGE OF EVALUATOR RESPONSES;
"D" INDICATES DISTRICTWIDE FIGURES; "S" SHOWS THIS SCHOOL ONLY.


COMPETENCY CATEGORIES ----->
PERSONAL QUALITIES
PROCEDURAL AND RECORD KEEPING SKILLS
EXPERTISE IN BASIC SKILLS AND SUBJECT AREAS
INSTRUCTIONAL SKILLS
CLASSROOM MANAGEMENT SKILLS
STUDENT/TEACHER RELATIONSHIPS
TOTAL
90%
80%
70%
60%
50%
40%
30%
20%
10%
PERCENT OF RESPONSES: 5 4 3 2 1
AISD : 20 46 34 1 0 | 10 34 55 1 0 | 16 43 41 1 0 | 12 40 47 1 0 | 17 43 39 1 0 | 19 48 32 1 0 | 16 43 40 1 0
SCHOOL: 0 28 71 0 0 | 0 3 97 0 0 | 2 18 81 0 0 | 0 10 89 1 0 | 0 22 78 0 0 | 0 21 79 0 0 | 0 18 81 0 0

NOTE: PERCENT OF RESPONSES FOR EACH COMPETENCY CATEGORY WAS COMPUTED AS FOLLOWS:
TOTAL RESPONSES OF A 1-5 RATING DIVIDED BY THE TOTAL RESPONSES OF ALL 1-5 RATINGS.

Figure 1: A Districtwide and School Summary of Teacher Evaluation Data

"in turn had higher ratings than did regular teachers. Elementary teachers and professionals almost always had higher ratings as a group than did their secondary counterparts. There were some shifts in whether male or female evaluators gave higher ratings and whether Anglo, Black, or Mexican-American evaluators gave higher ratings.

There is little doubt that this type of bias is endemic in evaluation ratings everywhere.

The Research Needs

After exploration of the realities of the teacher evaluation [illegible] it may well be asked if there is any real need to conduct research [illegible] a hopeless arena. However, the greatest reality of all is that just about every school system in the country has an evaluation system and most of them use rating scales. Indeed, the systems share features with very similar approaches used in industry and other governmental agencies. Whenever something has an almost universal existence, there is little point in arguing with it. Research, if it recognizes the basic realities, can only contribute to improvement.

An attempt has been made to outline the realities. What would the research that recognized these look like? At least some of the needs are described below.

1. Descriptive research of an indepth nature on current systems and practices is needed. Although this and other papers (Holley 1980) have made faltering attempts to give a picture of actual practice, they are but a sketchy glimpse of actual practice.

2. Experimental attempts to increase rater reliability must occur. The rating form is the most frequently used approach to teacher evaluation. Researchers have typically written it off as bad practice, but there is little evidence that reliability cannot be achieved. The authors suspect it may be possible. If so, the attempt should be made; research seems essential.

3. Data use is a big area requiring research. Many types of data can be considered; which of these data would enhance judgments? For example, would the collection and analysis of teacher tests and grades contribute significant knowledge about a teachers skills? What about lesson plans? Are student evaluations likely to contribute greatly?

4. Observation is the most typical evaluation technique used. What kind of observation training is most practical and yields the most benefits? How many hours are necessary for what levels of reliability?

It is the contention of this paper that there are great needs for teacher evaluation research. This research needs to be practical and rest on the realities of the world of practice, however. This paper has tried to at least sketch some of those realities as what it is hoped will be an encouragement for taking the needs seriously and as a challenge. Few fields of research can offer greater opportunities for impact on actual educational practice.

In summary then, the most urgent needs are for research on the design of systems for data collection and integration; viable methods of teacher observation that will increase observer reliability in practitioner settings (as opposed to research settings); the relationship between self-assessment and evaluator assessments; the incremental values of additional data sets such as student ratings; peer ratings, and supervisory ratings in the evaluation process; relationships between the various forms of ratings and student outcomes both learning and attitudinal.

References

Holley, F.M. An alternative approach to assuring teacher competence. Paper presented at the annual meeting of the American Educational Research Association, Boston, 1980.

Christner, C.A. The J. R. syndrome: Administrator bias in teacher evaluation. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, 1981.